

E 900788

NWC TP 7054

AD-A224 277

DTIC FILE COPY

Targeting Decisions Using Multiple Imaging Sensors: Operator Performance and Calibration

by
Marion P. Kibbe
and
Scott A. Weisgerber
Aircraft Weapons Integration Department

FEBRUARY 1990

NAVAL WEAPONS CENTER
CHINA LAKE, CA 93555-6001



Approved for public release; distribution is unlimited.

DTIC
ELECTE
JUL 12 1990
S B D

00 07 11 072

Naval Weapons Center

FOREWORD

This report documents two experiments that examined target identification performance and operator calibration using either a single information source or two separate sources. These experiments were performed at the Naval Weapons Center (NWC) during fiscal year 1989 under Task No. 62936N and 62234N, under Human Factors Block Funding and NWC Individual Exploratory Development Funding.

This report was reviewed for technical accuracy by Susan F. Savage and Steven Adams.

Approved by
E. K. KUTCHMA, Head
Aircraft Weapons Integration Department
25 January 1990

Under authority of
D. W. COOK
CAPT, USN
Commander

Released for publication by
W. B. PORTER
Technical Director

NWC Technical Publication 7054

Published by.....Technical Information Department
Collation.....Cover, 20 leaves
First printing.....140 copies

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE February 1990		3. REPORT TYPE AND DATES COVERED Summary; Sept 1988 - Dec 1989	
4. TITLE AND SUBTITLE Targeting Decisions Using Multiple Imaging Sensors: Operator Performance and Calibration				5. FUNDING NUMBERS	
6. AUTHOR(S) Marion P. Kibbe and Scott A. Weisgerber					
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Weapons Center China Lake, California 93555-6001				8. PERFORMING ORGANIZATION REPORT NUMBER NWC TP 7054	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Naval Ocean Systems Comand San Diego, California 92152-5000				10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES					
12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.				12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) Two experiments investigate target identification performance and operator calibration (i.e., the ability to evaluate the accuracy of one's own performance) using either a single information source or two separate sources. The experimental task required subjects to identify a target ship among two distractors using either simulated forward-looking infrared (FLIR) imagery, simulated range-only radar (ROR) imagery, or both information sources presented simultaneously. Relative to the single sensor condition, performance in the dual sensor condition could be either enhanced or decremented dependent upon the quality of the information presented on each sensor. In addition, both experienced pilot and non-pilot populations exhibited poor calibration, consistently underestimating the accuracy of their target identification performance. The finding that operators do not adopt optimal strategies for combining information from multiple sources suggests that performance could be enhanced by developing a set of integration rules. These rules would provide information regarding appropriate source weightings based on sensor image quality, and they would allow for the development of heuristics for information integration.					
14. SUBJECT TERMS Confidence judgment Decision making FLIR imagery Human factors of targeting Performance calibration Performance confidence ROR imagery Sensor integration performance Targeting performance				15. NUMBER OF PAGES 38	
				16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT SAR		

UNCLASSIFIED

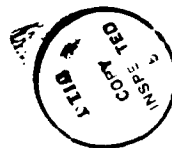
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

CONTENTS

Abstract.....	3
Introduction	3
Problem.....	4
Experiment One.....	5
Method.....	5
Results and Discussion.....	9
Experiment Two.....	18
Method.....	18
Results	18
Discussion.....	20
General Discussion.....	21
Performance.....	21
Calibration.....	21
Conclusions.....	22
Appendixes:	
A. Written Instructions to Subjects.....	23
B. Over- and Underconfidence, Calibration Scores, and Resolution Analysis.....	31
C. Information Integration Model.....	35
References.....	38



Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

ABSTRACT

Two experiments investigate target identification performance and operator calibration (i.e., the ability to evaluate the accuracy of one's own performance) using either a single information source or two separate sources. The experimental task required subjects to identify a target ship among two distractors using either simulated forward-looking infrared (FLIR) imagery, simulated range only radar (ROR) imagery, or both information sources presented simultaneously. Relative to the single sensor condition, performance in the dual sensor condition could be either enhanced or decremented dependent upon the quality of the information presented on each sensor. In addition, both experienced pilot and non-pilot populations exhibited poor calibration, consistently underestimating the accuracy of their target identification performance. The finding that operators do not adopt optimal strategies for combining information from multiple sources suggests that performance could be enhanced by developing a set of integration rules. These rules would provide information regarding appropriate source weightings based on sensor image quality, and they would allow for the development of heuristics for information integration.

INTRODUCTION

In order to facilitate all weather targeting at increased ranges, engineers are developing new imaging sensors, sensor suites, and autoclassifiers for use in a variety of Naval air and sea platforms. The operator of the future must evaluate and integrate multiple sources of information for target identification. These target identification decisions will typically be made under severe time constraints and in heavy work load environments where accurate target identification is essential to mission effectiveness.

This study is part of a series of experiments that focuses on operator targeting decisions using multiple sources of information. The first two experiments in this series (documented in References 1 and 2) evaluated methodological issues involved in determining if the accuracy of targeting decisions is a function of the number of information sources used in making the decision. The present experiment expanded upon the previous work by assessing targeting performance when two sources of sensor information were provided relative to targeting performance using a single source alone. This experiment also examined whether operator calibration (i.e., the

tendency to overestimate or underestimate the accuracy of one's own decisions) was present in these targeting decisions and if calibration was influenced by the number of information sources and/or the quality of the information.

PROBLEM

Decision Making. Aircrew targeting decisions will increasingly rely upon sensor and/or autoclassifier information presented in the cockpit rather than on close range visual inspection of the target. These targeting decisions will typically be based on information that is probabilistic or uncertain because current sensors rarely provide the level of detail required to identify a potential target with certainty. For example, forward-looking infrared (FLIR) and inverse synthetic aperture radar (ISAR) are two imaging sensors in which the quality of information on the display may vary with the atmospheric conditions, the range, and the aspect angle of the target. Unless everything is ideal, the images produced by these sensors will not be of sufficient quality to allow classification with complete accuracy and certainty. Similarly, autoclassifiers or automated target recognition systems provide only probabilistic classification of incoming sensor information. In general, autoclassifiers compare sensor output to an ideal image or to a set of target characteristics. These comparisons never yield exact matches so there is always some degree of uncertainty regarding target identity. Thus it can be seen that target identification based upon sensor or autoclassifier information is a decision based upon probabilistic and uncertain information.

It is possible that the inherent uncertainty associated with targeting sensor and autoclassifier outputs could be reduced by providing aviators with multiple sources of targeting information. As long as the information from multiple sensors is not precisely redundant, theories of information integration in most cases predict that the additional information provided by multiple sensors should result in superior target identification performance (Reference 1). But cogent arguments can also be made that multiple sensors may produce a performance deficit. For example, if the two sensors provide contradictory or ambiguous information, then the resulting conflict might lead to a degradation in performance. In addition the introduction of a second information source also increases the operator's processing demands, which in high work load situations might lead to a performance deficit.

Calibration. A large number of studies have found that decision makers tend to be inaccurate at assessing the quality of their decisions (References 3 and 4); this lack of calibration most often occurs in the form of overconfidence (i.e., decision makers think their performance is better than it actually is). This overconfidence has consistently been found in a variety of subject populations and with a wide variety of decision tasks. For example, college students are overconfident in deciding national

origin on the basis of children's art, in predicting stock market price movement on the basis of past performance, in sorting handwriting specimens on the basis of nationality, and in predicting their own performance on tests of general knowledge (Reference 3). Overconfidence can be lessened with training and by giving easier questions. Very easy items (i.e., items that 80% of the subjects answered correctly) sometimes lead to underconfidence, where the estimated level of performance is worse than the actual performance. Hard items tend to lead to overconfidence (Reference 3).

If poor calibration (whether overconfidence or underconfidence) is characteristic of decision makers who are uncertain about their decisions or who make predictions on the basis of uncertain information, then a lack of calibration should be present in targeting decisions. Furthermore, poor calibration could be magnified when targeting decisions are based on multiple sources of probabilistic and uncertain information. If makers of targeting decisions feel confident about poor decisions or if they are unsure of good decisions, then targeting performance would be adversely affected. To explore this possibility, both target identification performance and calibration of targeting decisions were studied as a function of the quality of the information presented in both the dual and single sensor cases.

EXPERIMENT ONE

This experiment compared targeting performance and decision calibration when target identifications were based on either single or dual sources of imaging target information. The two sensor sources were simulations of FLIR imagery and range-only radar (ROR) imagery. The images from these two sources were systematically varied in quality.

METHOD

Subjects. Twelve men and women professional employees of the Naval Weapons Center (NWC) served as subjects.

Materials

Range-Only Radar Images. Six ship images altered to simulate ROR were used as one source of imaging information. The ships, *Krivak*, *Kara*, *Sverdlov*, *Kashin*, *Kanin*, and *Iowa*, were taken from *Jane's Fighting Ships* (Reference 5). The

superstructures of the broadside images were outlined and digitized on a Genisco graphics processor using 60 evenly spaced points, and the image was then reduced so that each image was approximately 2.6 inches long. To simulate low, medium, and high levels of ROR distortion, each of these 60 profile points was altered vertically by adding or subtracting values randomly drawn from one of three distributions. These three distributions had a mean of zero and a standard deviation of five (low distortion condition), ten (medium distortion), or 20 (high distortion). Sixty new numbers were drawn from one of the distributions each time an image was shown on the screen so that the same image would never be shown twice.

Forward-Looking Infrared Images. To simulate FLIR images, the same six ships were photographed broadside from *Jane's Fighting Ships* (Reference 5) and digitized using an Imaging Technology Digital Image Processor. These digitized images were equated for size in terms of pixel count by altering the hulls of the ships without altering the superstructures. The images were then reduced so that each ship measured roughly 1.5 centimeters at the waterline. The final ship images were white and appeared on a light gray background.

As was the case with the ROR images, there were three levels of FLIR distortion. Two separate distortion techniques were used. First, a 9 by 9 filter mask was passed over all profiles except those in the low distortion condition. This mask acted as a low pass filter that blurred the edges of the profile, the blur increasing with the number of passes. The filter mask was not used in the low distortion condition; the medium level of distortion used two passes of the filter; and the high distortion level used four passes. Second, so that the same image would not be seen repeatedly (and to add distortion to the low distortion condition) random noise masks were superimposed on each image. Twenty different variants of each level of distortion (0, 2, or 4 passes of the filter) were created for each of the six ships.

There was no effort to equate the low, medium, and high levels of distortion between the simulated FLIR and ROR images. Similarly, the distortion levels were not matched to a particular level of ship identification performance or to any real sensor parameters, such as range or atmospheric conditions. On the basis of the previous experiments, it was assumed that better targeting performance would be associated with better image quality (References 1 and 2).

Equipment. A VAX 11/750 computer controlled the presentation of stimuli and the recording of data. The VAX controlled a Panasonic optical disk recorder (Model TQ-2023F) on which the FLIR images were recorded and a Genisco graphics processor that generated the ROR profiles. The FLIR and ROR images were presented on two 9.5-inch Satchel Carlson 10M915 cathode ray tubes (CRTs). The subject interface was a Texas Instruments (TI) portable professional computer, which was also

controlled by the VAX. The voice capability of the TI computer was used in the training session for feedback.

Tasks

Single Display. In the single source task, either three FLIR images or three ROR images were presented on one of the two CRTs. The task was a three-alternative, forced choice recognition task in which the subject was asked to select which of the three ships was the *Krivak*. The *Krivak* was always the target. Non-targets were counterbalanced combinations of two ships selected from five distractor ships. The position of the *Krivak* with respect to the non-targets on the CRT was systematically varied. The subject responded by pressing the 1, 2, or 3 key on the TI keyboard to designate which of the three ships was the *Krivak*. There were eight trials at each distortion level and a total of 24 trials for each sensor in the single display conditions.

In addition to the designation of the target, subjects were asked to provide their confidence rating for each selection. Confidence judgments could range from 30 to 100% in increments of 10. Because only ratings evenly divisible by 10 were accepted, subjects were informed that 30 represented a chance level of responding and 100 represented complete certainty. Complete instructions are given in Appendix A.

Dual Display. In the dual sensor task, both FLIR and ROR imagery were presented simultaneously using both CRTs. As before, the target ship was always the *Krivak*. The target and distractor ships were in the same order on both sensors and were presented at the same level of distortion within each sensor. Between sensors however, the distortion level varied. Over all of the trials in the experiment, each of the three levels of distortion on one sensor was paired with each of the three levels of distortion for the other sensor, yielding nine possible combinations of distortion levels. These combinations ranged from both sensors having low levels of distortion to both sensors having high levels of distortion. Eight trials were presented at each of the nine pairings, for a total dual-display block of 72 trials. Subjects stated their level of confidence after each targeting decision.

Procedure. Subjects were tested on two consecutive days. On the first day, training was provided to distinguish the *Krivak* from the distractor images for both the FLIR and ROR profiles; instruction was also given for making confidence judgments. Subjects then practiced, with performance feedback for each trial on each of the three tasks: FLIR decisions alone, ROR decisions alone, and dual-display decisions. On the second day, with no further training and no performance feedback, a test of each of the three sensor conditions was given.

Training and Practice Session

Forward-Looking Infrared. A printed version of an ideal (i.e., nondistorted) FLIR image of the *Krivak* was shown to the subject and the most salient features distinguishing the *Krivak* FLIR image from the distractors were described. Then 15 FLIR trials (five at each distortion level) were shown on the CRT. On each trial, the *Krivak* was identified so that the subject could study the characteristics of the image. A practice session then followed in which subjects were asked to identify the *Krivak* and to state their confidence in each judgment. The practice session consisted of 24 trials, with target position and distractor ship identity systematically varied. Feedback was given after each response so that subjects could monitor their performance. The feedback indicated the accuracy of the response and also gave the correct identification following an error.

Range-Only Radar. In a similar manner, subjects were trained to recognize a ROR image of the *Krivak*. A printed version of the non-distorted outline was shown, and the most salient characteristics were described. Nine different versions of the *Krivak*, along with distractor ships (three at each distortion level), were shown to the subjects in hard-copy form. The *Krivak* was always identified so that subjects could compare it to the distractor ships. Following training there was a practice session of 24 trials on the CRT (eight at each distortion level), which included feedback. As described above, target position and distractor ship identities were systematically varied. As with FLIR practice, subjects identified which ship was the *Krivak* and stated their level of confidence in each decision.

Dual Practice. Finally, subjects were given practice using both information sources together. Seventy-two trials were presented, eight at each of the nine possible dual sensor distortion levels. In the dual source practice, as with the single source, subjects had to identify which of the three ships was the *Krivak* and to state their level of confidence in the choice. Again, feedback indicating both the accuracy of the response and the correct choice following an error was presented after each trial.

The training and practice session lasted approximately one hour, but the actual times varied as there were no time constraints placed on any of the decisions.

Test Session. Testing occurred on the following day. In the testing session, subjects were presented with each of the three tasks (i.e., single display FLIR, single display ROR, and dual display). A printed version of the ideal FLIR and ROR images of the *Krivak* was provided as a reference. The task order was counterbalanced across subjects. On each trial, subjects were to designate which of the three ships was the *Krivak* and rate their level of confidence in that decision. Feedback on performance was not provided during the test trials. Testing took approximately 40 minutes, but as with

the practice session, the length of time varied as there were no constraints imposed on the time allowed for making the targeting decisions or confidence judgments.

RESULTS AND DISCUSSION

Scoring. The first session was considered practice and was not analyzed. The data from the second session were scored in terms of the percentage of correct identifications and the mean confidence rating at each level of distortion. The difference between the performance and confidence scores was also analyzed.

An additional set of difference scores was also calculated in order to compare performance and confidence changes between the single and dual display conditions. First considering only performance, two difference scores were calculated: one compared dual display performance to performance on the single FLIR display and the other compared the dual display performance to performance on the single ROR display. To calculate the dual - FLIR (dual minus FLIR) performance scores, the FLIR performance score for each distortion level in the single sensor condition was subtracted from each of the three performance scores associated with that FLIR distortion level in the dual display case. Thus there were nine dual - FLIR difference scores for each performance. A positive score indicates that dual display performance was higher than single display performance, while a negative score indicates that single display performance was superior. Similarly, nine dual - ROR performance scores were also calculated.

The same set of difference scores was also calculated for the confidence ratings by subtracting single FLIR confidence and single ROR confidence scores from the dual display confidence scores at each distortion level. Scores above zero indicate an increase in confidence in the dual display case, while scores below zero indicate a confidence decrease for dual displays.

Single Source Analyses. Target identification performance, decision confidence, and the difference between performance and confidence were analyzed using separate analyses of variance. These analyses used a factorial, repeated measures design with two completely crossed within-subject factors: information source (with two levels, FLIR and ROR); and distortion (with three levels, low, medium, and high). The dependent measure in the first analysis was the percentage of correct identification of the *Krivak*; the dependent variable in the second analysis was the average confidence rating for the decisions; and the dependent variable in the third analysis was the difference between the performance and confidence measures. The results of these analyses are shown in Figure 1.

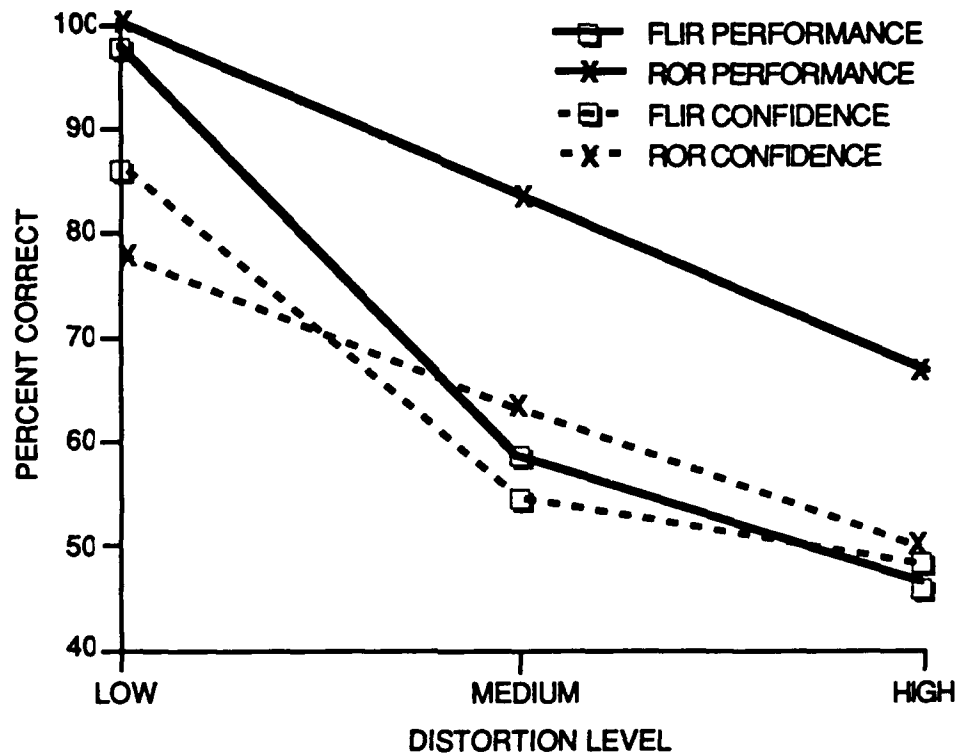


FIGURE 1. Single Sensor Performance and Confidence as a Function of Distortion Level.

The analysis of performance yielded a significant effect of information source, $F(1, 11) = 34.01$, $p < 0.0001$; and distortion, $F(2, 22) = 77.76$, $p < 0.0001$; and a significant interaction between information source and distortion, $F(2, 22) = 4.89$, $p < 0.05$. As can be seen in Figure 1, performance on both FLIR and ROR images decreased with increases in distortion level, but even at the highest level of distortion, targeting judgments remained well above the 33% (chance) level of performance. Target identification performance with the simulated FLIR images was consistently worse than with ROR images and the performance decrease attributable to increasing the level of distortion from low to medium was more rapid for the FLIR images than for the ROR images. This is probably due to the fact that the distortion levels for the two sensors were not equated for difficulty, and the quality of the FLIR images appeared to degrade more rapidly than the quality of the ROR images.

The analysis of the calibration data indicated that the effect of distortion level was significant, $F(2, 22) = 86.88$, $p < 0.0001$, as was the information source x distortion level interaction, $F(2, 22) = 4.41$, $p < 0.025$. However, the effect of information source was not significant, $F(1, 11) < 1.0$, indicating that calibration in targeting decisions did not vary as a function of whether FLIR or ROR imagery was used. As

was the case with performance, confidence scores decreased with increases in distortion. That is, as the quality of information coming from either source decreased, targeting performance became worse and subjects were correspondingly less confident in the targeting decisions that they made. The calibration scores closely paralleled the performance scores, but were generally lower.

The analyses of the difference scores found that performance scores were significantly higher than the confidence scores, $F(1, 11) = 12.76$, $p < 0.01$, indicating that subjects were underconfident of their ability to correctly identify the target ship. There was also a significant effect of information source, $F(1, 11) = 10.02$, $p < 0.01$, reflecting the fact that the difference between performance and confidence (i.e., the underconfidence) was larger for ROR than for FLIR.

Traditionally, calibration research using confidence judgments groups together all decisions that had the same confidence rating (e.g., all decisions at the 60% confidence level). Comparisons are then made between the actual performance in each group and the confidence ratings. This approach is shown in Appendix B. In this analysis, the subjects were also shown to be consistently underconfident.

Dual Source Analyses. Performance, confidence, and the difference between performance and confidence were analyzed using three separate analyses of variance. Each analysis employed a two-way, fully crossed, within subjects design with ROR distortion as the first factor (with three levels of distortion) and FLIR distortion as the second factor (also with three levels of distortion). In the first analysis the dependent measure was targeting performance, in the second the dependent variable was decision confidence, and the third analyzed the difference between performance and confidence. These results are shown in Figure 2 and explained below.

The analyses of the performance scores showed significant main effects and interactions for all factors: for ROR distortion, $F(2, 22) = 61.05$, $p < 0.0001$; for FLIR distortion, $F(2, 22) = 22.36$, $p < 0.0001$; and for ROR distortion x FLIR distortion, $F(4, 44) = 4.89$, $p < 0.01$. Similarly, the confidence scores yielded significant effect of ROR distortion, $F(2, 22) = 64.64$, $p < 0.0001$; FLIR distortion, $F(2, 22) = 22.36$, $p < 0.0001$; and the ROR distortion x FLIR distortion interaction, $F(4, 44) = 5.98$, $p < 0.001$. These results are shown in Figure 2. Comparison of performance and confidence indicated that the confidence scores were lower than the performance scores, $F(1, 11) = 11.75$, $p < 0.01$. There was also a significant effect of ROR distortion, $F(2, 22) = 4.12$, $p < 0.05$, reflecting the fact that the underconfidence was larger at the lower ROR distortion levels (see Figure 2). The effect of FLIR distortion and the ROR x FLIR interaction were not significant.

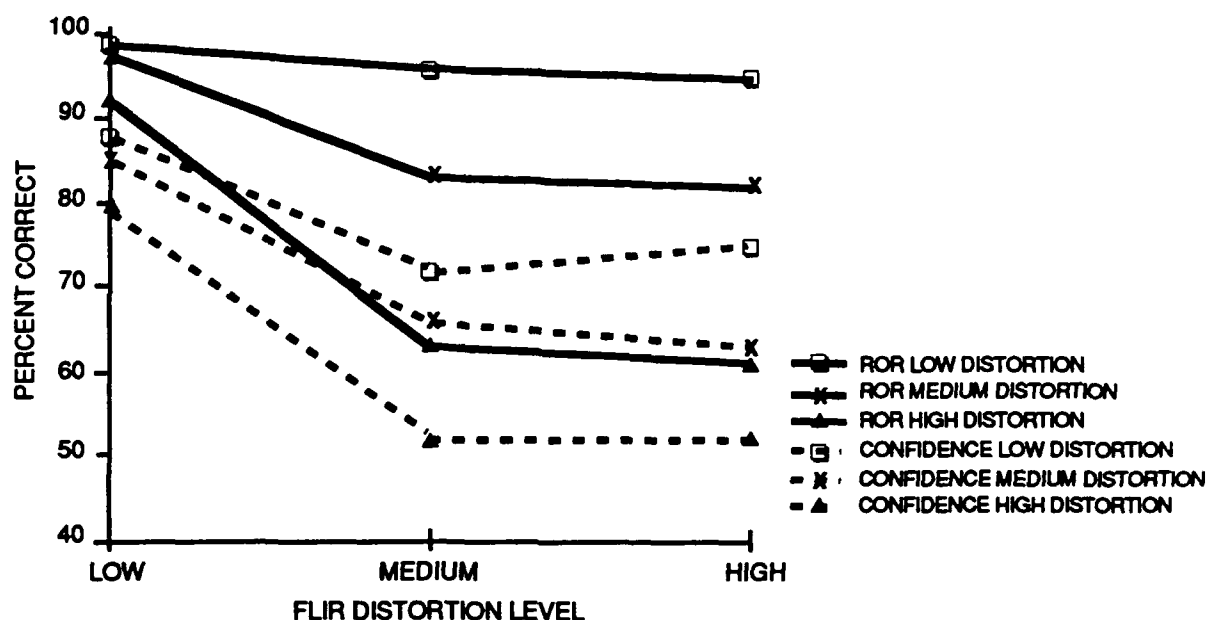


FIGURE 2. Dual Sensor Performance and Confidence as a Function of Combined Distortion Levels.

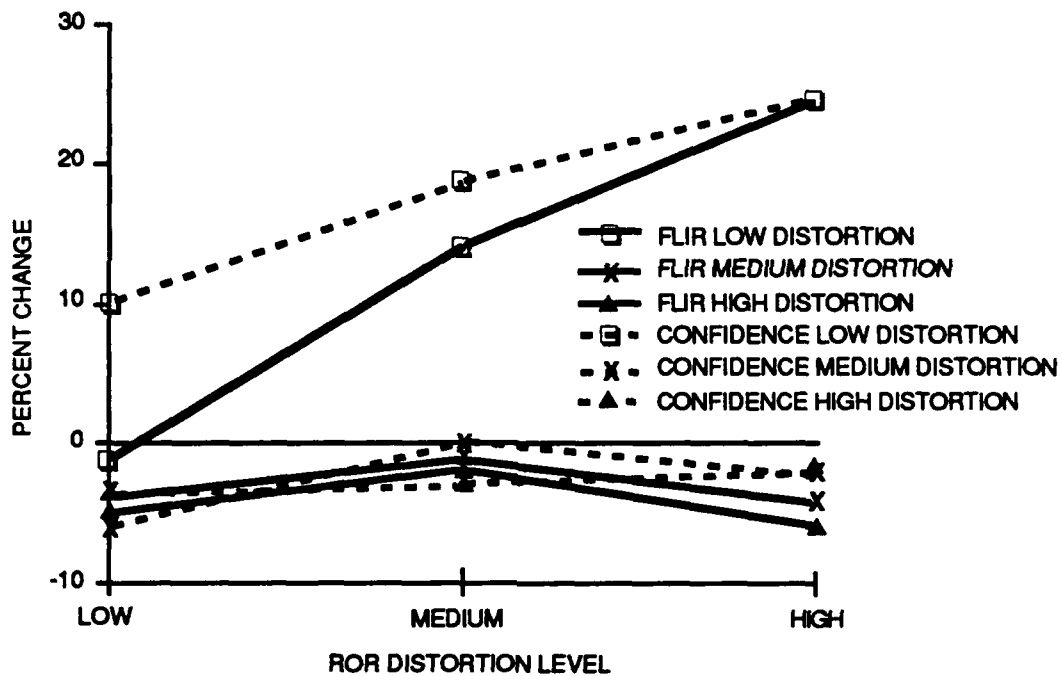
Examination of the performance scores in Figure 2 shows that when either the ROR or the FLIR imagery was presented at the lowest level of distortion, identifications of the *Krivak* were extremely accurate (correct more than 90% of the time). When at least one of the sensors gave a very good image (low distortion), subjects apparently used it to make the identification and ignored the second sensor source that was of lesser quality. However, when both sensors were at either the medium or high level of distortion, the effect of the distortion varied between the two sensors. Performance dropped regularly with each increase in ROR distortion. For FLIR however, performance decreased from low to medium distortion, but there was no further decrease between the medium and high distortion levels. The finding that performance was more sensitive to ROR than to FLIR distortion suggests that either the medium and high levels of FLIR distortion did not differ or that when the FLIR imagery was distorted, subjects relied primarily on the ROR image to make the identification. The latter interpretation is supported by the fact that in the single sensor condition, performance decreased between the medium and high levels of FLIR distortion, $F(1, 11) = 6.57, p < 0.05$.

Overall, subjects continued to be able to make surprisingly accurate judgments even with very distorted information. At the highest level of both FLIR and ROR distortion the *Krivak* was still correctly identified over 60% of the time. As with the single sensor data, the confidence scores followed the same pattern as the performance scores, except that the confidence scores were consistently lower. Once again, the subjects underestimated their performance. Decision confidence decreased markedly when a very good image was paired with a distorted image, even though performance was largely unaffected and remained close to 100%. This indicates that subjects were sensitive to the quality of the imagery on the second sensor even though it did not effect performance. As with performance, confidence decreased with each increase in ROR distortion but was insensitive to the difference between the medium and high levels of FLIR distortion.

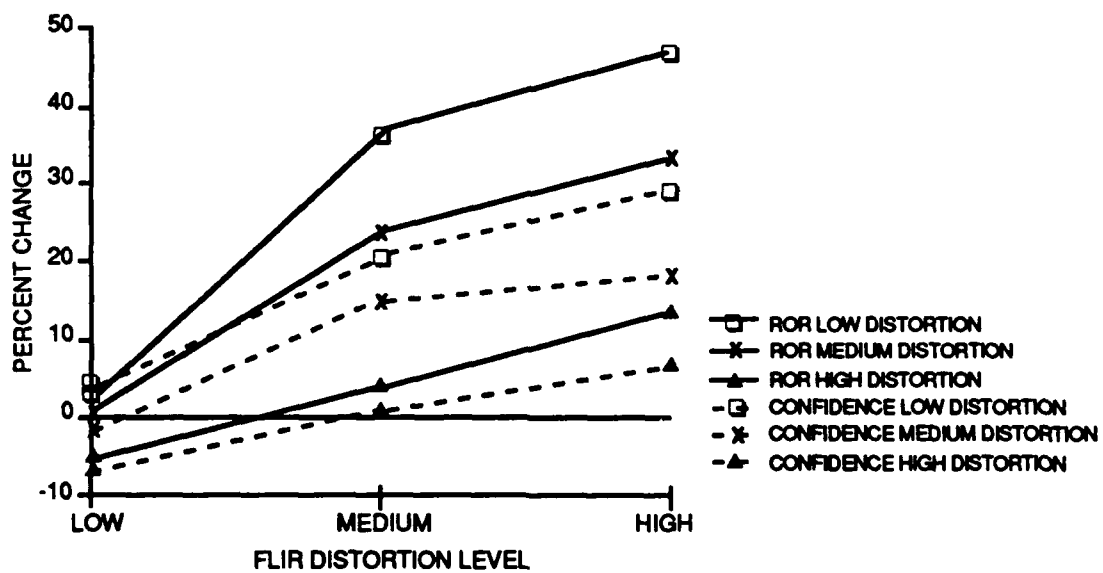
Dual Source Compared to Single Source. These analyses compared targeting performance and calibration using a single source of information with performance and calibration using two sources of information. Four, fully-crossed, repeated-measures designs were analyzed using two-way analyses of variance (ANOVAs). The first factor in each analysis was FLIR distortion with three levels (low, medium, and high), and the second factor was ROR distortion, also with three levels (low, medium, and high). The dependent measures in the ANOVAs were the four difference scores described in the scoring section (dual – ROR performance, dual – FLIR performance, dual – ROR confidence, and dual – FLIR confidence).

The ANOVAs using the two performance difference scores, dual – ROR and dual – FLIR, yielded quite divergent results. For dual – ROR, the grand mean was not significantly different from zero, and the main effect of ROR was also not significant. The main effect for FLIR, $F(2, 22) = 22.36$, $p < 0.0001$, was statistically significant, as was the interaction of ROR x FLIR, $F(2, 22) = 4.89$, $p < 0.01$. These results are shown in Figure 3a. For the analysis of the dual – FLIR difference score, however, the grand mean, all main effects, and interactions were significant; the grand mean, $F(1, 11) = 21.81$, $p < 0.001$; ROR main effect, $F(2, 22) = 61.05$, $p < 0.0001$; FLIR main effect, $F(2, 22) = 16.7$, $p < 0.0001$; ROR x FLIR interaction, $F(2, 22) = 4.89$, $p < 0.01$. These results are shown in Figure 3b.

The grand mean, all main effects, and interactions for both dual – ROR and dual – FLIR confidence scores were significant. The results of these analyses are given in Table 1 and shown in Figures 3a and 3b.



(a) Dual - ROR performance change and confidence change as a function of combined distortion levels.



(b) Dual - FLIR performance change and confidence change as a function of combined distortion levels.

FIGURE 3. Analyses Results.

TABLE 1. Dual - ROR and Dual - FLIR Calibration Scores Significant Results.

Source of variation	df	F	Significance
Dual - ROR analysis			
Grand mean	1, 11	6.31	<0.03
ROR	2, 22	3.82	<0.04
FLIR	2, 22	79.47	<0.0001
ROR x FLIR	4, 44	5.92	<0.0007
Dual - FLIR analysis			
Grand mean	1, 11	11.86	<0.0055
ROR	2, 22	67.29	<0.0001
FLIR	2, 22	27.60	<0.00001
ROR x FLIR	4, 44	5.72	<.0009

Figure 3a shows the dual sensor performance scores minus the performance scores with the ROR source alone. Another way of describing this figure is that it shows the performance gained (or lost) by adding FLIR to the performance based on ROR alone. Points above zero on the figure indicate performance enhancement; points below zero indicate performance deficit. The figure shows that adding a very good FLIR image to ROR either (1) leaves performance unchanged (if ROR distortion was low) or (2) enhances performance (if the ROR image was moderately or highly distorted). The addition of a distorted FLIR image to any level of ROR image does not improve performance above that based on ROR alone--in fact it leads either to no performance change or to a performance decrement. This lack of enhancement for distorted FLIR is reflected in the ANOVA by the fact that the grand mean is not significantly different from zero. The interpretation of decision confidence difference scores follows the interpretation of performance results exactly and does not offer any new insights.

Figure 3b shows the performance improvement when ROR information is added to the performance that was based on FLIR imagery alone. When good or moderately distorted ROR is added to good FLIR images, there is no performance change, but when they are added to moderately or very distorted FLIR, performance is enhanced. Performance is also enhanced when very distorted ROR is added to distorted FLIR, but very distorted RORs added to very good FLIR leads to a slight performance deficit. Apparently people do not ignore the poor ROR images; they try to integrate the poor

RORs with the good FLIRs and they alter decisions that ideally should have been based on the very good FLIRs alone. Once again, confidence difference scores are very similar to performance difference scores. The addition of ROR to FLIR does seem to augment confidence in targeting decisions in almost all instances when the FLIR is moderately or severely distorted.

Reaction Times. Subjects were given no instructions concerning time limits, and they were allowed all the time that they wanted to study each display before reaching a decision. Reaction times (RT) were recorded for each targeting decision for both single and dual sensor presentations. A one way analysis of variance comparing reaction times for dual sensor, FLIR alone, and ROR alone showed a significant difference between the mean reaction times, $F(2, 22) = 8.43$, $p < 0.01$. The mean RT for using two sensors was 13.91 seconds, while for ROR alone it was 10.49 seconds and for FLIR alone it was 10.41 seconds. Even though there was a significant increase in RT in the dual sensor condition, it is clear that less time was devoted to studying each sensor in the dual case than in the single case. Reaction times were also examined as a function of distortion. In the single sensor conditions the effect of sensor type was not significant, $F < 1$, indicating that subjects did not spend different amounts of time using the different imagery sources. There was a significant effect of distortion, $F(2, 22) = 9.61$, $p < 0.001$, and a significant sensor type by distortion interaction. Examination of the RTs in Table 2 indicates that when using ROR, RT increased with increased distortion, while with the FLIR, more time was spent on the medium level of distortion than on the low and high levels. In the dual sensor condition, there was a significant main effect for ROR distortion, $F(2, 22) = 45.73$, $p < 0.01$, and a significant interaction between ROR x FLIR distortion, $F(4, 44) = 3.32$, $p < 0.05$. There was no significant main effect for FLIR distortion.

TABLE 2. Reaction Times for Single Sensors by Distortion Level.

Distortion level	RT, s
ROR	
Low	7.01
Medium	10.76
High	13.71
FLIR	
Low	8.47
Medium	12.45
High	9.79

Mean RTs for ROR increased from 10.76 seconds at the low distortion levels to 17.47 seconds at high distortion levels. There was no corresponding increase in RT for FLIR. Examination of the RTs in Table 3 shows that in the nine combinations of distortion levels for the two sensors, RTs were largely dependent on the distortion level of the ROR. The exception to this rule, which explains the significant interaction, is that the RTs were shorter when distorted RORs were paired with very good FLIRs. The reaction time analysis confirms the findings from the dual – ROR analysis: people do not integrate degraded FLIR information when ROR information is available. RT with single FLIR of medium distortion is longer than with all other FLIR, but as Figure 1 shows, people given enough time can use the FLIR information at medium distortion.

TABLE 3. Dual Sensor Reaction Times for the Nine Combinations of Distortion Levels.

ROR distortion level	FLIR distortion level	RT, s
Low	Low	10.69
Low	Medium	11.74
Low	High	9.85
Medium	Low	13.55
Medium	Medium	13.09
Medium	High	13.81
High	Low	14.80
High	Medium	18.79
High	High	18.85

Information Integration Model. The dual sensor performance data from this experiment were compared to the decision combination model referenced by Foyle (Reference 1). This analysis characterizes dual sensor operator performance as "enhanced," "super-enhanced," or "decremented" as compared to single sensor performance. A fourth category, "failed integration," was added to characterize dual sensor performance that fell between the best and worse single sensor performances (see Appendix C for a complete discussion of the Information Integration Model analysis). The nine distortion combinations for dual sensor performance were characterized using these four categories. There was no evidence of better integration performance at any particular combination of the distortion levels. Single subject analyses using the model showed that there were large individual differences in subjects' ability to integrate information from multiple sources; some subjects showed enhancement or super-enhancement in every distortion level, while others were rarely able to improve their performance.

EXPERIMENT TWO

The results of Experiment 1 indicated that subjects consistently underestimated the accuracy of their target identification performance. This finding was surprising given that most studies have reported that subjects are highly overconfident of their performance in decision making tasks. The population tested in the first experiment consisted of NWC scientists and engineers. It is possible that these results would not generalize to the population of pilots and Naval Flight Officers (NFOs), who are the ultimate users of multisensor targeting systems. Pilots and NFOs have generally had experience in making targeting decisions (although not with the present simulated sensor imagery), are trained to make rapid and firm decisions, and thus may tend to have a different decision calibration than the scientists and engineers. Therefore, it was decided to test a small group of pilots and NFOs to determine whether they would show a similar pattern of underconfidence.

METHOD

The method used in this experiment was identical to that reported in Experiment 1. Six Naval reserve pilots and NFOs served as subjects in the experiment.

RESULTS

The primary concern of this study was to determine whether pilots exhibited a similar pattern of underconfidence to that found in the non-pilot population. In the single sensor condition, analysis of the difference between the performance and confidence scores indicated that pilots significantly underestimated their targeting performance, $F(1, 5) = 13.07$, $p < 0.025$ (see Figure 4). The effects of sensor type, distortion level, and the interaction were all nonsignificant.

Analysis of the performance scores indicated that the effect of sensor type was marginally significant, $F(1, 5) = 6.45$, $0.05 < p < 0.01$, reflecting a tendency toward worse performance with FLIR than with ROR. The effect of distortion level was highly significant, $F(2, 10) = 20.35$, $p < 0.001$, indicating that performance decreased with increases in distortion. There was no evidence of a sensor type by distortion interaction, $F < 1$. In terms of confidence, both sensor type, $F(1, 5) = 11.75$, $p < 0.05$, and distortion level, $F(2, 10) = 62.00$, $p < 0.0001$, were

significant, as was the interaction, $F(2, 10) = 10.85$, $p < 0.01$. Examination of Figure 4 indicates that confidence was lower for the FLIR trials than for the ROR trials and that confidence decreased more rapidly with increases in distortion for FLIR than for ROR.

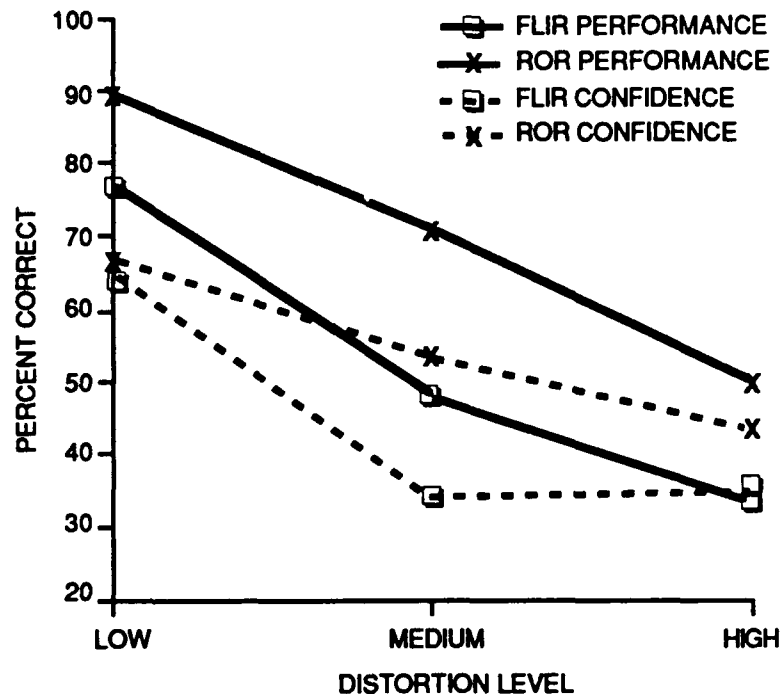


FIGURE 4. Single Sensor Performance and Confidence as a Function of Distortion Level.

Analysis of the difference between performance and confidence in the dual sensor condition also indicated that the pilots underestimated their performance even when both sensors were available, $F(1, 5) = 262.97$, $p < 0.0001$. There was also a significant effect of ROR distortion level, $F(2, 10) = 5.96$, $p < 0.05$, reflecting the fact that pilot lack of calibration was greatest at the low distortion levels where performance was best (see Figure 5). Pilots were underconfident in a similar pattern for the FLIR distortion levels but the effect did not reach significance, $F(2, 10) = 3.5$, $0.05 < p < 0.1$. There was no evidence of a ROR by FLIR interaction, $F < 1$.

Examination of the performance scores indicated that performance decreased with increases in ROR distortion, $F(2, 10) = 51.70$, $p < 0.0001$, while there was no evidence that FLIR distortion affected performance, $F(2, 10) = 2.86$, $p > 0.1$, and no interaction, $F(4, 40) = 2.28$, $p > 0.05$ (see Figure 5). Confidence was influenced by

both ROR distortion level, $F(2, 10) = 21.08$, $p < 0.001$, and by FLIR distortion level, $F(2, 10) = 21.40$, $p < 0.001$. There was no evidence of a ROR by FLIR interaction, $F < 1.0$.

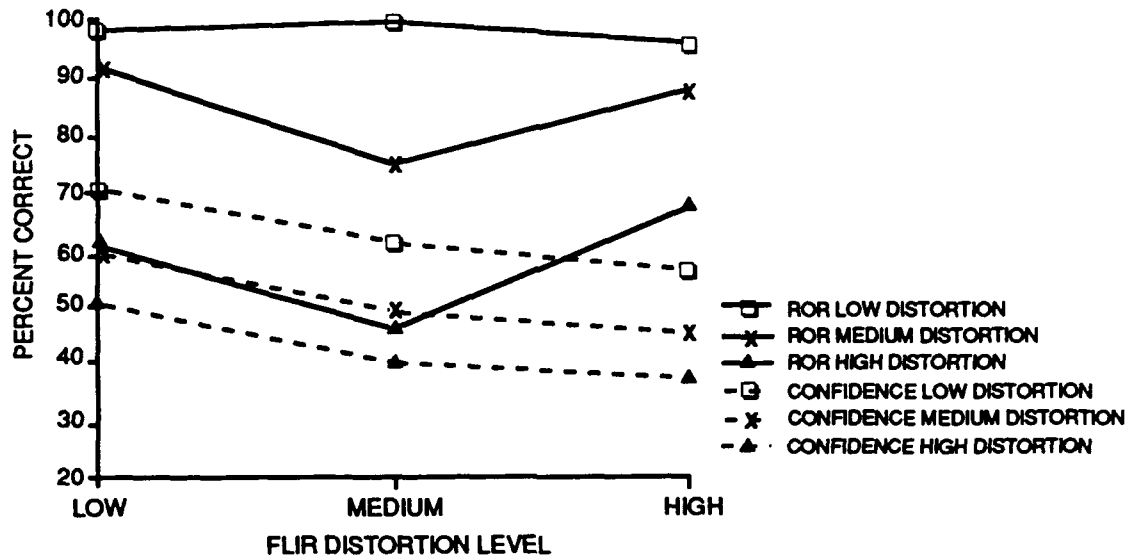


FIGURE 5. Dual Sensor Performance and Confidence as a Function of Combined Distortion Levels.

Due to the small sample size used in the present experiment, the more detailed analyses comparing single and dual sensor performance and confidence were not conducted.

DISCUSSION

The results of Experiment 2 are essentially the same as those found in Experiment 1. The pilots and NFOs used in this study lacked calibration and underestimated their ability to make accurate targeting decisions in both the dual and the single sensor conditions. In addition, as was the case in the first experiment, performance and calibration were better with simulated ROR imagery than with simulated FLIR imagery. The fact that there was not a significant decrease in performance attributable to FLIR distortion level suggests that the pilots relied more on the ROR than on the FLIR. However, pilot estimates of performance did decrease as a function of FLIR distortion level, which indicates that the subjects were sensitive to the quality of the information on the second sensor even though it did not affect performance.

GENERAL DISCUSSION

PERFORMANCE

The results of this study demonstrate that in using dual sensor imaging displays, if either of the sensors provides high quality (i.e., relatively undistorted) information then target identifications are extremely accurate and the quality of the imagery on the other sensor has a minimal impact on performance. However, when the information on either sensor is more distorted, there is an interaction between the quality of the information and the type of sensor. In these situations it is clear that humans do not optimally combine information from multiple sources. In situations where the quality of the information on one of the sensors was worse than that provided by the other sensor, performance may be worse than it would have been if only the better of the two sensors had been presented.

The fact that the sensors in the present study were simulations of actual sensors severely limits any conclusions that can be drawn regarding the actual sensors. For example, the fact that performance with the simulated ROR imagery tended to be better than performance with the simulated FLIR imagery could have been a product of the simulations themselves.

CALIBRATION

The results of this study have shown that in both experiments operators consistently underestimated the accuracy of their target identification performance. Underconfidence was a characteristic of both experienced aviators and NWC employees. This finding is in contrast with previous studies, which have reported that people tend to be highly overconfident when predicting their accuracy levels in most decision making tasks. One situation where underconfidence has been reported is when the tasks are very easy and performance is highly accurate (80% correct, see Reference 3). In the experiments documented here, performance was highly accurate (above the 80%) when either of the imagery sources was at the minimal distortion level, and this may have led subjects to be underconfident. However, subjects were also underconfident when using highly distorted imagery, which was associated with performances of 40 to 50%, suggesting that subjects making targeting decisions would be underconfident regardless of their level of performance.

A second factor that has been shown to improve calibration is training (Reference 3). It is possible that subjects in the present experiments did not receive sufficient training in target identification using FLIR and ROR imagery. However, the fact that the overall level of performance was very accurate in both experiments suggests that this was not the case. In addition, as a pretest to the present experiments, four expert subjects who each had over 20 hours of experience with the target identification task were evaluated to determine their confidence levels. Their data closely paralleled the results of Experiments 1 and 2 and showed the same pattern of underconfidence. Therefore, it seems unlikely that underconfidence in these studies is based upon a lack of practice in making targeting decisions with this type of sensor imagery.

CONCLUSIONS

This study has demonstrated the importance of evaluating human capabilities and limitations in the integration of information from multiple sources. The fact that operators do not always make effective use of multiple sources of information suggests that it may not always be advisable to provide the operator with all of the available information that exists in a multisensor suite. For example, if the operators tend to overweight poor quality information, then it is possible that this information should either not be presented or be presented with a caveat noting that the quality of the information is poor. Further experiments will determine whether subjects can use information on image quality as a basis for sensor integration and thus improve their targeting performance. It seems likely that the tendency to overweight poor quality information in integrating information from multiple sources would hold regardless of the sensor sources that were used. A similar finding has often been reported in investigations of multiple cue probability learning (Reference 4, pp. 12-13). However, an investigation of the specific tradeoffs that exist between information quality and sensor type would require that the experiment be repeated using real sensor imagery.

The finding that operators tend to underestimate the accuracy of their targeting decisions also has important implications for the design of targeting systems. If pilots underestimate their ability to make accurate targeting decisions on the basis of imaging sensor information there may be resultant critical delays in weapons release decisions. The possibility that inaccurate calibration degrades targeting performance indicates that further research is needed concerning mechanisms for improving calibration.

NWC TP 7054

Appendix A

WRITTEN INSTRUCTIONS TO SUBJECTS

PRACTICE SESSION

INTRODUCTION

This experiment studies the accuracy and confidence of targeting decisions based on information which varies in quality. You will be shown simulations of two sensors, FLIR and ROR, and the information on each will vary in quality. Initially, you will be shown each sensor simulation separately and then finally you will see the two sensors together each showing the same targets.

The equipment that will be used in this experiment is the TI computer and the two displays in front of you, and the laser disk recorder to your right. Will you please read and then sign these consent forms so that you can participate in this experiment? Page one of the forms verifies that you have had the experiment and its equipment described to you, and page two describes how we will use the data collected from the experiment.

Training

To participate in this experiment you will need to learn how ships (particularly our target ship, the Krivak) look on FLIR and ROR. You will learn first about FLIR.

This picture shows the silhouette of the Krivak. A broadside FLIR image of a ship looks something like a small blurred picture of it. In the experiment, the image will be white on a gray background. The bottom picture of the Krivak might be something like an ideal FLIR image but is less blurry than a real FLIR image would be. We can use this image however as a starting point for learning to recognize and select the Krivak image from among a set of simulated FLIR images of other ships.

It is difficult to describe the characteristics that distinguish the simulated FLIR image of the Krivak. First, notice that the superstructure of the ship becomes grouped together and appears as one blurred "hill" right in the center of the ship. Some people notice some extra blurriness just aft of this hill.

In front and behind the "hill," the ship looks very small and flat all the way to the bow and the stern. Some people note a small bump at the stern which at low levels of distortion helps to discriminate the Krivak from other ships.

At higher levels of distortion in this simulation, the whole ship seems to blur out evenly. That is, even though the center is larger than the rest, it does not appear brighter than either the bow or stern. Finally, when seen at high levels of distortion in this simulation, the Krivak image often seems to be more faded than most of the images of other ships.

The drawings I am about to show you are simulations of ROR. This first drawing looks something like a smoothed over outline of the silhouette of the Krivak, shown here on the bottom of the picture. This drawing is the best, most ideal ROR profile of the set; the other drawings in the set are distortions of this one, simulating noise or interference in the sensor reception. The aim of this training session is to teach you to be able to recognize the Krivak ROR profile, and be able to pick it out from a set of ROR profiles of other ships.

Let's first examine this "ideal" ROR simulation. It's most prominent feature is the "V" in the center of the profile. It is important to notice that the "V" is about equidistant from the bow and the stern. Generally the front side of the "V" is a bit taller than the stern side. The front side is also thicker, and comes down toward the deck in a zig-zag pattern. All of these clues will help you to distinguish the Krivak from other ships.

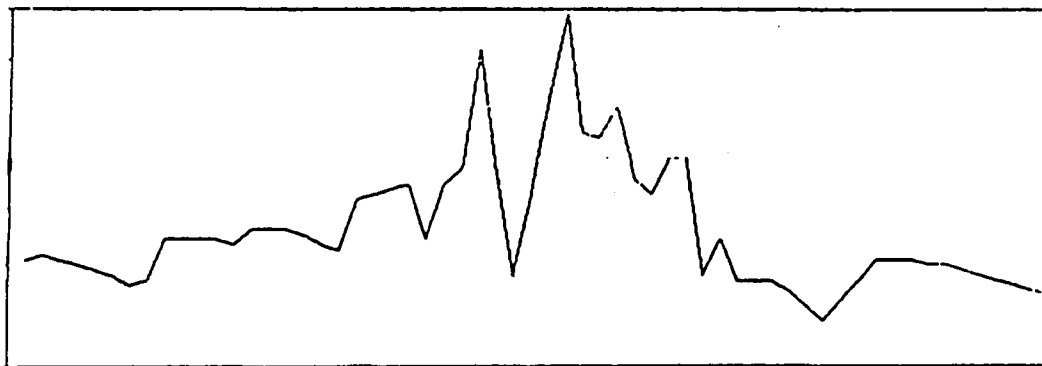
Next notice the "mound" or "bump" in the bow of the profile. Notice that there is a dip which precedes it, and that the "bump" is almost as wide as the front part of the "V". Learn to look for this "bump" as a characteristic of the Krivak.

Now we will look through the other drawings. In these drawings, there is more distortion or noise in each drawing, but many of the characteristics described above can still be seen at least to some degree. The Krivak is always in these drawings the top ship on the left. In each drawing, look for:

1. The "V"
2. Its location in the center
3. Its height
4. The thickness of the forward arm of the "V"
5. The bump in the bow
6. The dip before the bump
7. The relative size of the bump

NWC TP 7054





target

The Krivak Decision

In all cases you will see simulated images of three ships, and your problem will be to select which one is the "Krivak". Keep the ideal images of the Krivak for reference as you make your decision. When you have decided which ship is the Krivak, to indicate your choice type "1" if it is the left most FLIR, "2" if it is the center ship, and "3" if it is the right ship. For ROR images, each image is shown in a different, numbered quadrant, and you need to type in the quadrant number of your selection.

Confidence Judgment

After each Krivak selection, we would like to know how confident you feel that this was a good decision, so you will next be asked, "What is your percentage confidence rating?" Here we want to know your judgment that if you were given this same quality of information many different times, what percent of the time would you make a correct decision? 30%? 40%? 70%? If you think that the Krivak could have been any of the three ships that were shown on the screen, that is, that the information was so bad that you simply had to guess, then you should type in 30% (sheer chance would be 33%, but only percents divisible evenly by ten are acceptable in this experiment). If you are quite sure that one of the three images is not the Krivak, but it could be either of the other two and you have no idea which, then type in 50%, as you are guessing between the two. If however you are doing better than guessing, if the information on the screen allows you to make a better than chance judgment, then the percent you type in should be higher than either 30 or 50. Perhaps you are quite sure that it is ship number 1: you might give a confidence of 80; or 90; or if you are completely certain, 100. You may choose any of the following levels of certainty for each decision: 30, 40, 50, 60, 70, 80, 90, 100. You need not type in the percent sign, only the number.

Overall Rating

You will have three blocks of practice, and three blocks of testing on the following day. The first two practice blocks will show only one sensor, either FLIR alone or ROR alone. The third will show you both FLIR and ROR, and the same ships will be shown for each sensor in the same order. At the end of each block, please use the rating scale next to you on the table. Give your initials, whether the block was FLIR, ROR, or both, and your guess of your overall score in percents.

Summary

In summary, for each screen or set of two screens: first type in the number answering "Which is the Krivak?" (1, 2, or 3). Next type in the number between 30 and 100 which best represents your confidence rating. Finally, for each block, give your overall rating of your percent correct. Thank you.

NWC TP 7054

Appendix B

**OVER- AND UNDERCONFIDENCE, CALIBRATION
SCORES, AND RESOLUTION ANALYSIS**

Lichtenstein and Fischhoff used three separate measures to evaluate the accuracy of subjective confidence ratings (Reference 1). For the purposes of the present paper these measures will only be described in general terms; the computational formulas can be found in Reference 1. The first measure is over- or underconfidence, which is defined as the difference between the overall means of the confidence and performance scores. The second is calibration, which is a measure of the absolute value of the difference between the performance and confidence scores. Both over- and underconfidence contribute equally to the larger (i.e., less accurate) calibration score. The final measure is resolution, which deals with both the granularity and the accuracy of the confidence ratings (i.e., the ability of the subject to accurately assign different levels of confidence to match performance). A high resolution score indicates that confidence ratings fall into categories that maximally separate the confidence scores from the mean level of performance. Individuals who use only a limited number of confidence ratings (e.g., 50% or 100%) will tend to have low resolution scores.

A calibration curve for the present study can be seen in Figure B-1, which plots the actual mean percent correct for each of the confidence rating categories. It can be seen that in almost every instance subjects were underconfident. Performance was higher than confidence for every rating except 100%, where it is not possible to be underconfident. The overall mean performance score was 85.1% and the mean confidence score was 69.1%. The resulting confidence score of -16.0 indicates that on the average, subjects were underconfident by 16%. The calibration score was 16.8, indicating that on the average, the absolute value of the confidence ratings differed from the performance scores by 16.8%. The fact that the calibration and confidence scores were almost the same reflects the fact that the scores were consistently underconfident. The analysis of resolution yielded a low score of 0.014, indicating that there was not a great deal of separation between the confidence ratings and the mean level of performance.

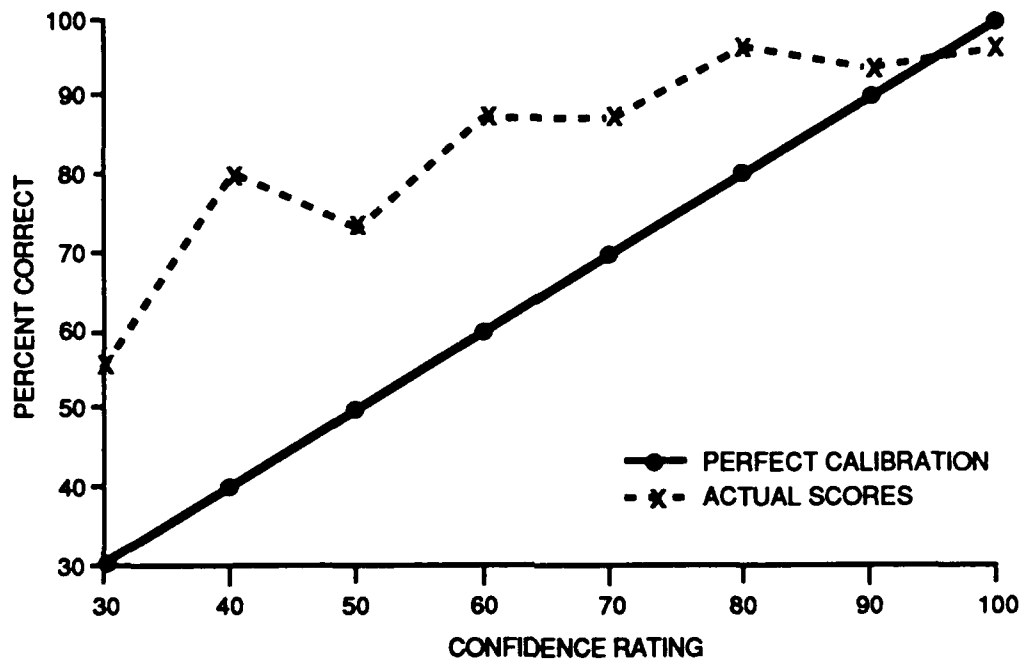


FIGURE B-1. Percent Correct by Confidence Rating.

NWC TP 7054

Appendix C
INFORMATION INTEGRATION MODEL

The dual sensor performance data from this experiment were compared to the decision combination model referenced by Foyle (Reference 1). The model states "that performance with a complex stimulus is predictable from the performance with the individual stimuli according to the following equation:

$$P_{12} = P_1 + P_2 - P_1 P_2$$

where p_1 and p_2 represent detection probabilities for the two stimuli presented in isolation and p_{12} is the detection probability when both stimuli are available," Foyle characterizes dual sensor performance in his report as "enhanced" if it equals or exceeds the highest single sensor performance, "super-enhanced" if it equals or exceeds the performance computed from the decision combination model, and "decremented" if it is less than the highest single sensor performance.

In the analysis documented here, the same "enhanced" and "super-enhanced" categories were used to characterize performance, but in addition, the "decrement" category was subdivided into two sections: "failed integration" when performance with two sensors falls between the performances for each of the two single sensors; and "decrement," which is performance at or below the lowest single sensor. Figure C-1 shows the frequency for each of the four characterizations of performance when levels of distortion for the dual sensors are taken into account.

When the distortion level of either of the sensors is low, performance is near 100% for both dual and the single sensor conditions and, therefore, there is no possibility that subject performance exceeds the performance calculated by the model (super-enhancement). Therefore, in the first five distortion conditions shown in Figure C-1, there was no observed super-enhancement; most performances were characterized as "enhanced." In the four conditions that use combinations of medium and high distortion, dual sensor performance for the twelve subjects fell almost evenly into all four categories. Thus better integration did not seem to be favored by any particular combination of information quality.

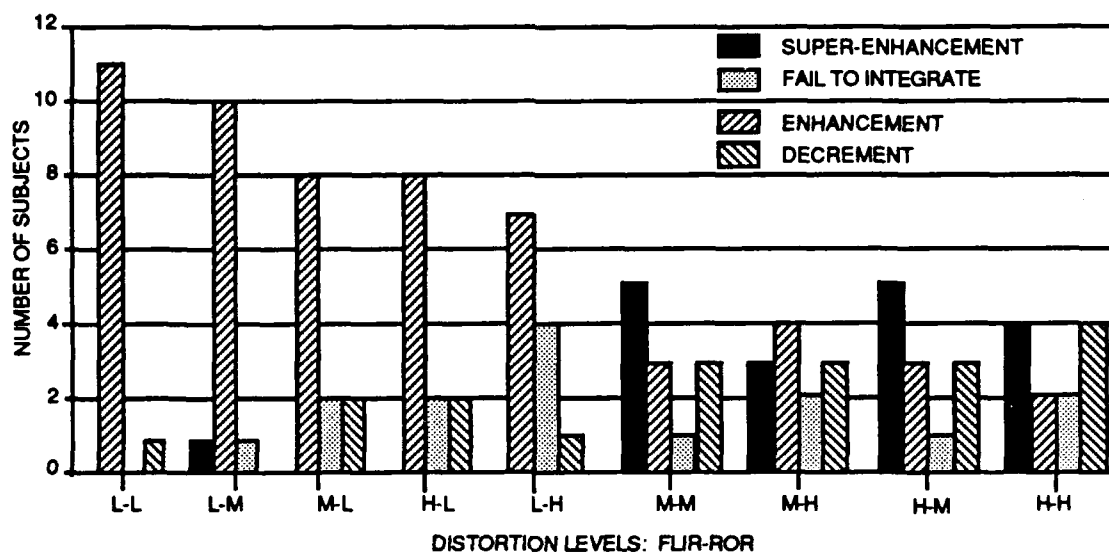


FIGURE C-1. Performance Analysis Integration Model Characterizations.

Looking at single subject performance (Table C-1), the performance of three subjects (no. 4, 5, and 6) could be characterized as showing enhanced or super-enhanced performance for all distortion levels. At the other extreme, one subject (no. 12) showed enhanced performance for only three of the nine types of decisions; five of the remaining six of his decisions were characterized as "decrement," and one showed "failed integration." Because there was little difference between this subject and those at the other extreme in single sensor performance, there apparently are large individual differences in the ability to integrate information from dual sensors and make good targeting decisions from that information.

TABLE C-1. Single Subject Analysis, Integration Model.

Distortion levels for FLIR/ROR	Subjects											
	1	2	3	4	5	6	7	8	9	10	11	12
Low/low	E	E	D	E	E	E	E	E	E	E	E	E
Medium/low	D	E	E	E	E	E	E	F	E	F	E	D
High/low	D	E	D	E	E	E	E	E	E	E	F	F
Low/Medium	E	E	E	E	SE	E	E	E	F	E	E	E
Medium/medium	F	D	SE	E	E	E	SE	SE	SE	SE	D	D
High/medium	F	D	E	E	SE	E	SE	F	SE	SE	SE	D
Low/high	F	E	F	E	E	E	E	F	D	E	F	E
Medium/high	SE	D	E	E	SE	E	F	D	F	SE	E	D
High/high	SE	D	E	SE	SE	SE	D	F	F	D	E	D

E = enhanced; D = decremented; F = failed integration; SE = super-enhanced.

REFERENCES

1. Naval Weapons Center. *Multisensor Evaluation Framework*, by David C. Foyle. China Lake, Calif., NWC, September 1989. (NWC TP 7027, publication UNCLASSIFIED.)
2. ----- . *Multisensor Integration Using Two Independent Sources*, by Susan F. Savage. China Lake, Calif., NWC, August 1989. (NWC TP 6983, publication UNCLASSIFIED.)
3. S. Lichtenstein and B. Fischhoff. "Do Those Who Know More Also Know More About How Much They Know?" *Organizational Behavior and Human Performance*, Vol. 20 (1977). Pp. 159-83.
4. P. Slovic, B. Fischhoff, and S. Lichenstein. "Behavioral Decision Theory." *Annual Review of Psychology*, Vol. 28 (1977). Pp. 1-39.
5. *Jane's Fighting Ships*, Royal Navy, Capt. John Moore, ed. London, England, Jane's Publishing Company, 1982.